Supplementary Note

Detailed comparison of an African MDR *S*. Enteritidis with P125109

Malawian invasive strain D7795 was selected for detailed analysis. D7795 was isolated from the blood of a Malawian child in 2000, and is MDR (resistant to amoxicillin, chloramphenicol, cotrimoxazole and tetracycline). Unsurprisingly, the draft genome sequence of D7795 closely resembles that of P125109. It does, however harbour a number of predicted pseudogenes, a different prophage repertoire and a substantially larger plasmid with an MDR region. In total D7795 shares 95.2% of its coding sequences with P125109 and P125109 shares 97.0% of its CDS with D7795. There are a further 363 genes in D7795 which have NS-SNPs with respect to P125109.

Evidence of recent gene acquisition in P125109 has previously been highlighted; of the 130 CDSs specific to P125109, compared to *S.* Gallinarum strain 287/91, those associated with the novel genomic region denoted region of difference (ROD) 4 and prophages ΦSE10, ΦSE14 and ΦSE20 appear to be recent acquisitions with no evidence of them ever being present in *S.* Gallinarum strain 287/91. The same is partially true of D7795, which contains no remnant or genomic 'scar' to suggest the two neighbouring variable regions from P125109, ROD21 or ΦSE20 were ever present. It does, however contain the degraded remnant of a prophage Gifsy 2-like region, ΦSE12, although this region is even more degraded than in P125109. In contrast, P125109 contains the scar or remnant of a different Gifsy 2-like prophage region, which appears complete in D7795. These signature events are consistent with D7795 and P125109 having a common ancestor but forming distinct clades.

The core genome of P125109 is largely conserved in D7795. All 12 SPIs and all 13 fimbrial operons associated with P125109 are present. The genes within the SPIs are also intact, however fimbrial protein Y in the *fim* operon is a pseudogene, and 7 other fimbrial operon genes contain non-synonymous SNPs. The fimbrial operons

are otherwise intact with respect to P125109 and D7795 has maintained the multiple functional fimbrial operons common to host promiscuous Salmonellae[1]. These differences are put in detailed context in Supplementary Table 5, adapted from Thomson et al[2].

The principal difference between the two chromosomes is in the prophage regions and genomic islands (RODs). P125019 contains prophage ΦSE20 and near to it, genomic island ROD 21. Prophage region ΦSE20 encodes 51 CDSes in ~41 kbp. Potential cargo genes including the type-three secretion-system (TTSS) secreted effector protein *sseK3* and fragments of other TTSS effector genes including *sopE*, *sspH2* and *gogA*[2]. It has recently been demonstrated in both a mouse model[3] and cell-line model[4] that genes associated with region ΦSE20 are likely to be necessary for invasion in mice and chickens. The function of ROD21 is less clear. It is the largest genomic island in P125109 (~ 26 kb) and its structure has previously been described in detail[2]. Of note, it has a low G+C content and most of this island encodes a paralogue of H-NS (*hnsB* – a down regulator of transcription) and/or an H-NS antagonist, *hnsT* [5-7].

In addition to these large differences, some of the prophage regions and RODs have undergone further degradation involving specific important genes. ΦSE12 and ΦSE12A are prophage remnants lying adjacent to each other and carry remnants of the PhoPQ-activated genes *pagK* and *pagM*, and an intact copy of *sopE* known to stimulate cytoskeletal re-organisation during invasion of host cells. This region has undergone further degradation in D7795, which does not carry *sodCI*, encoding a Cu/Zn superoxide dismutase known to be an important colonisation factor for *S.* Typhimurium[8]. Genomic island ROD 30 has lost the nickel/cobalt transporter *rcnA*. Both of these regions are likely to be ancient acquisitions as they are present in *S*. Typhimurium LT2.

D7795 contains three complete prophage regions that are absent from P125109, containing a total of 150 predicted CDS regions, in addition to one Gifsy 2-like

region, highly degraded in P125109, but retaining a further 10 genes in D7795. The three complete regions were annotated using PHAST (Phage Search Tool: http://phast.wishartlab.com/) and are closely related to phages Fels-2 (33.1 kb, 47 predicted CDS, including 3 cargo genes), Enterobacter P88 (38.1 kb, 52 CDS, including 13 cargo genes) and a Gifsy 1-like prophage (33.1 kb, 52 CDS, including 5 cargo genes). The regions largely contain hypothetical genes or essential phage machinery. However the Enterobacter P88-like prophage carries a gene predicted to encode a leucine rich repeat (LRR) protein. These LRR-domains are protein interaction regions and have been implicated in suppression of gut-inflammation through inhibition of NF-κB-dependent gene expression by both *S*. Typhimurium and *Shigellae* [9].

Inspection of the draft assembly of D7795 reveals 42 putative pseudogenes. Whilst this only represents 0.1% of the genome, a striking feature is how many are involved in core metabolic pathways, specifically in genes critical for an enteric lifestyle, and how many genes from the same pathway contain NS-SNPs. Some of them have catabolic roles in colonic mucus degradation, including SEN1434 (hexonate metabolism) and the sulphatase SEN0035. Colonic mucus contains sugar acids, a source of nutrients that gut dwelling organisms must be able to utilize [10]. Several of the apparent pseudogenes are involved in anaerobic metabolism, including SEN3902, a glycerol dehydrogenase. Glycerol metabolism is necessary for anaerobic growth of enterobacteria on non-fermentable sources. Dimethyl sulphoxide reductase, a gene necessary for the use of terminal electron acceptors in anaerobic metabolism, has also been lost, together with two other sulphur metabolism genes. Two genes involved in formate metabolism have become degenerate. There are a further four metabolic pseudogenes, including mono-pentose pathway (L-fuculose kinase [*fucK*]*)*.

Gene *pocR* is a pseudogene in D7795. This is one of the most important regulatory genes for the B12 synthesis and propanediol utilisation pathway[11], a key enteric-adaptation pathway. Many *Enterobacteriaceae* have lost the capacity to synthesise

cobalamine and therefore to degrade 1,2-propanediol. However, *Salmonella* re-acquired a 40 kb metabolic island encoding both the *cbi* and *pdu* loci[12-14]. 1,2-propanediol is an important source of energy for *S.* Typhimurium, especially within the intracellular compartment [15]. Of note, the B12 transport gene *yncD* is also a pseudogene and a number of downstream genes in this pathway have acquired NS-SNPs including two tetrathionate and three propanediol utilisation genes, which suggest the possibility of functional changes of this pathway. Similar changes in this pathway are also seen in invasive, host-adapted serotype *S*. Gallinarum[2].

Like *S.* Typhi and *S.* Gallinarum 287/91, *S*. Enteritidis D7795 carries a number of mutations in a gene which is centrally involved with shedding - *shdA*. This gene is a surface-localized, fibronectin-binding protein whose expression is induced in vivo in the murine caecum[16,17] and is carried on a 25-kb genetic island named centisome 54 (CS54 island) in *S.* Typhimurium. Absence of this gene is associated with reduced faecal carriage and shedding of *S*. Typhimurium in mice, but not pigs[16,18]. This gene is also a pseudogene in *S.* Typhi and *S.* Gallinarum and whilst it is intact in the African invasive *S*. Typhimurium ST313, the related gene *ratB* from CS54 has become a pseudogene[2,19]. As discussed above, there is one fimbrial gene (*fimY*), which is a pseudogene and one type III secretion system protein.

In addition to the metabolic signal suggestive of host adaptation, there are seven pseudogenes amongst membrane transport genes, including the two further genes involved in cobalamine metabolism. Cobalamine synthase incorporates Ni-Fe and Cobalt respectively, and it is interesting to note that D7795 has lost *rncA*, a high affinity Ni/Co transporter, through a deletion in ROD30 (see above)[20] and the gene encoding the cobalamine transporter *yncB*. Five other putative, uncharacterised transport genes are pseudogenes. A further seven membrane-protein associated genes have mutations.

Three DNA replication and repair genes carry mutations, including a *mutT*-family gene. Only one pseudogene lies in an amino acid catabolic or biosynthetic pathway,

L-serine (*L*-serine and *L*-threonine catabolism), for which there is redundancy[21]. A further five hypothetical genes and one phage-protein gene have become pseudogenes. Supplementary Table 3 provides a list of some of the common traits identified amongst the functions of genes lost independently by D7795, *S.* Typhi and *S*. Gallinarum. Some of the overlapping traits are striking; including the loss of genes involved in common metabolic processes such as cobalamine and propanediol utilisation and electron transport acceptor function.

In addition to the large differences in terms of the variable region of the genome, there is further evidence of reductive evolution in D7795 in the form of 363 genes containing NS-SNPS. The disproportionate clustering around membrane structures is yet another signal seen in host adaptation with both *S*. Typhi[22] and *S*. Gallinarum[2]. Many of these genes will be functional, but it is likely that some have become pseudogenes and that others have altered functions. Certainly there is evidence of clustering around pathways, for example the cobalamine (b12) biosynthetic pathway.

The MDR virulence plasmid of the Southern African *S*. Enteritidis clade

Aligned of pSENT-BT against finished P125109 plasmid pSENV revealed evidence of two deletions; the first contained genes *srgA, srgB and luxR*, part of the *lux* operon and the second was within gene pSENV0033, a putative integrase. All the other genes on pSENV including the virulence operon, were intact. The additional regions aligned against two sequenced fragments of an *S.* Enteritidis virulence plasmid (pUO-SeVR) isolated from an African patient presenting with MDR iNTS Enteritidis in Spain[23]. In total, nine antimicrobial resistance genes were identified, encoding resistance to amoxicillin (*bla*TEM-1), tetracycline (*tetR* and *tetA*), chloramphenicol (*cat*A2) the components of cotrimoxazole (*dfr*A7, *sul*1 and *sul*2) and aminoglycosides (*str*A and *str*B).

The plasmid contained a number of additional genes that might be associated with virulence and a toxin/antitoxin plasmid addiction system. These included *tir*, associated with gut wall attachment in enteropathogenic *E. coli*, and associated with virulence in 0157:H7 strain[24]; a *pecM* gene which is a membrane spanning efflux protein which may have drug efflux properties[25]; and *pncA* which is a nicotinamidase whose function is unclear, however it is essential to the pathogenesis of *Borrelia burgdorferi*[26]. Lastly genes *mucA* and *mucB* are present. The latter has been shown to be a lesion bypass polymerase and although their precise role is unclear, together they may repair damaged plasmid DNA, or facilitate mutagenesis. This would be beneficial for an MGE transmitted through a number of bacterial hosts, by allowing faster adaptation to foreign intracellular environments[27].

Like pSENV, the virulence plasmid contained an incomplete set of conjugal transfer genes (*tra* genes) and it was therefore predicted that this plasmid would not be capable of conjugal transfer. This was tested by conjugation experiments at 26°C and 37°C using the plasmid from *S*. Typhimurium A54560[28] and IncHI1 plasmid PHCM1 as positive controls and PT4-like *S*. Enteritidis A1636 as a negative control. As predicted, the plasmids of A1636 and D7795 failed to transfer, whilst the positive control successfully transferred.

Interpretation of high-throughput phenotyping

The findings from the pairs of replicates were compared using principal component analysis and found to be highly consistent. In total, there were 200 statistically significant differences between log-fold change in adjusted signal values (SVs) between the 2 isolates; 97 at 28°C and 103 at 37°C. Following manual review of the signal from each well, 80 of the results were considered genuinely significant and pursued further. The entire output is graphically displayed in Figure 3. Evaluation using Pathway Tools software to identify genes involved in each specific pathway with evidence of increased or decreased metabolic activity revealed that 14/27

(52%) of pathways with evidence of decreased metabolic activity at 28°C in D7795 also displayed evidence of genomic degradation, and 12/30 (40%) of pathways with evidence of decreased metabolic activity at 37°C in D7795 also displayed evidence of genomic degradation.

The two replicates of D7795 showed enhanced metabolism of three simple carbohydrates, including glucose, sucrose and lactulose, also of D-saccharic acid, D-alanine, mucic (galactaric) acid and formic acid. There were far more instances of reduced activity, many of which have a corresponding signature of genomic degradation, such as pseudogene formation or the presence of NS-SNPs in genes corresponding to the pathway. These included dulcitol and glycolic acid in the glycerol degradation pathway, propionic acid in the propanediol pathway and ethylamine and ethynolamine. Also there was reduced activity in response to three forms of butyric acid, a short chain fatty acid. This is another signature of loss of enteric adaptation, as the ability to metabolize short chain fatty acids is an extremely useful trait for enteric adaptation[29]. The full list of differences is detailed in Supplementary Table 4 and 5.

There was no increased metabolism of any nitrogen sources in D7795 compared to A1636, however metabolism was reduced in several nitrogen sources (Supplementary Table 4 and 5). The most interesting sources noted to be down regulated were alloxan and allantoic acid. Allantoin can be found in the serum of birds, and is utilised as a carbon source during *S.* Enteritidis infection of chickens[30] and pseudogenes relating to allantoin have been noted in *S.* Typhimurium ST313[19]. D7795 showed increased metabolism in the presence of 6% sodium chloride solution at both temperatures, but decreased metabolism in the presence of 4% sodium formate, which is consistent with the numerous mutations in formate catabolizing genes in D7795.

References

1. Townsend, S.M. *et al.* Salmonella enterica serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect Immun* **69**, 2894-901 (2001).
2. Thomson, N.R. *et al.* Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res* **18**, 1624-37 (2008).
3. Silva, C.A. *et al.* Infection of mice by Salmonella enterica serovar Enteritidis involves additional genes that are absent in the genome of serovar Typhimurium. *Infect Immun* **80**, 839-49 (2012).
4. Shah, D.H., Zhou, X., Kim, H.Y., Call, D.R. & Guard, J. Transposon mutagenesis of Salmonella enterica serovar Enteritidis identifies genes that contribute to invasiveness in human and chicken cells and survival in egg albumen. *Infect Immun* **80**, 4203-15 (2012).
5. Park, H.S. *et al.* Induction chemotherapy with idarubicin plus N4-behenoyl-1-beta-D-arabinofuranosylcytosine in acute myelogenous leukemia: a newly designed induction regimen--a prospective, cooperative multicenter study. *Semin Hematol* **33**, 24-9 (1996).
6. Navarre, W.W. *et al.* Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. *Science* **313**, 236-8 (2006).
7. Williamson, H.S. & Free, A. A truncated H-NS-like protein from enteropathogenic Escherichia coli acts as an H-NS antagonist. *Mol Microbiol* **55**, 808-27 (2005).
8. Figueroa-Bossi, N., Uzzau, S., Maloriol, D. & Bossi, L. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella. *Mol Microbiol* **39**, 260-71 (2001).
9. Haraga, A. & Miller, S.I. A Salmonella enterica serovar typhimurium translocated leucine-rich repeat effector protein inhibits NF-kappa B-dependent gene expression. *Infect Immun* **71**, 4052-8 (2003).
10. Sweeney, N.J., Laux, D.C. & Cohen, P.S. Escherichia coli F-18 and E. coli K-12 eda mutants do not colonize the streptomycin-treated mouse large intestine. *Infect Immun* **64**, 3504-11 (1996).
11. Bobik, T.A., Ailion, M. & Roth, J.R. A single regulatory gene integrates control of vitamin B12 synthesis and propanediol degradation. *J Bacteriol* **174**, 2253-66 (1992).
12. Lawrence, J.G. & Roth, J.R. Evolution of coenzyme B12 synthesis among enteric bacteria: evidence for loss and reacquisition of a multigene complex. *Genetics* **142**, 11-24. (1996).
13. Porwollik, S., Wong, R.M. & McClelland, M. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. *Proc Natl Acad Sci U S A* **99**, 8956-8961. (2002).
14. Prentice, M.B. *et al.* Cobalamin synthesis in Yersinia enterocolitica 8081. Functional aspects of a putative metabolic island. *Adv Exp Med Biol* **529**, 43-6 (2003).

15. Klumpp, J. & Fuchs, T.M. Identification of novel genes in genomic islands that contribute to Salmonella typhimurium replication in macrophages. *Microbiology* **153**, 1207-20 (2007).
16. Kingsley, R.A., Santos, R.L., Keestra, A.M., Adams, L.G. & Baumler, A.J. Salmonella enterica serotype Typhimurium ShdA is an outer membrane fibronectin-binding protein that is expressed in the intestine. *Mol Microbiol* **43**, 895-905 (2002).
17. Kingsley, R.A. *et al.* Molecular and phenotypic analysis of the CS54 island of Salmonella enterica serotype typhimurium: identification of intestinal colonization and persistence determinants. *Infect Immun* **71**, 629-40 (2003).
18. Boyen, F. *et al.* The fibronectin binding protein ShdA is not a prerequisite for long term faecal shedding of Salmonella typhimurium in pigs. *Vet Microbiol* **115**, 284-90 (2006).
19. Kingsley, R.A. *et al.* Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res* **19**, 2279-87 (2009).
20. Rodrigue, A., Effantin, G. & Mandrand-Berthelot, M.A. Identification of rcnA (yohM), a nickel and cobalt resistance gene in Escherichia coli. *J Bacteriol* **187**, 2912-6 (2005).
21. Sawers, G. The anaerobic degradation of L-serine and L-threonine in enterobacteria: networks of pathways and regulatory signals. *Arch Microbiol* **171**, 1-5 (1998).
22. Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* **413**, 848-52 (2001).
23. Rodriguez, I., Rodicio, M.R., Guerra, B. & Hopkins, K.L. Potential international spread of multidrug-resistant invasive Salmonella enterica serovar enteritidis. *Emerg Infect Dis* **18**, 1173-6 (2012).
24. Bono, J.L. *et al.* Association of Escherichia coli O157:H7 tir polymorphisms with human infection. *BMC Infect Dis* **7**, 98 (2007).
25. Rouanet, C. & Nasser, W. The PecM protein of the phytopathogenic bacterium Erwinia chrysanthemi, membrane topology and possible involvement in the efflux of the blue pigment indigoidine. *J Mol Microbiol Biotechnol* **3**, 309-18 (2001).
26. Jewett, M.W., Jain, S., Linowski, A.K., Sarkar, A. & Rosa, P.A. Molecular characterization of the Borrelia burgdorferi in vivo-essential protein PncA. *Microbiology* **157**, 2831-40 (2011).
27. Goldsmith, M., Sarov-Blat, L. & Livneh, Z. Plasmid-encoded MucB protein is a DNA polymerase (pol RI) specialized for lesion bypass in the presence of MucA', RecA, and SSB. *Proc Natl Acad Sci U S A* **97**, 11227-31 (2000).
28. Feasey, N.A. *et al.* Drug Resistance in Salmonella enterica ser. Typhimurium Bloodstream Infection, Malawi. *Emerg Infect Dis* **20**, 1957-9 (2014).
29. Keeney, K.M. & Finlay, B.B. Enteric pathogen exploitation of the microbiota-generated nutrient environment of the gut. *Curr Opin Microbiol* **14**, 92-8 (2011).
30. Dhawi, A.A. *et al.* Adaptation to the chicken intestine in Salmonella Enteritidis PT4 studied by transcriptional analysis. *Vet Microbiol* **153**, 198-204 (2011).